(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
13 June 2002 (13.06.2002)

PCT

(10) International Publication Number
WO 02/46467 A2

(51) International Patent Classification[7]: C12Q 1/68

(21) International Application Number: PCT/IB01/02811

(22) International Filing Date: 7 December 2001 (07.12.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/254,090	8 December 2000 (08.12.2000)	US
10/007,926	7 December 2001 (07.12.2001)	US

(71) Applicant *(for all designated States except US)*: **IPSOGEN** [FR/FR]; 232 Boulevard Sainte-Marguerite, F-13273 Marseille Cedex 09 (FR).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **BERTUCCI, François** [FR/FR]; Le Sully, Parc de la Cadenelle, 122, rue du Commandant Rolland, F-13008 Marseille (FR). **HOULGATTE, Rémi** [FR/FR]; 31, Boulevard Lombard, F-13015 Marseille (FR). **BIRNBAUM, Daniel** [FR/FR]; 9, rue Baldacchini, La Croix-du-Sud, F-13009 Marseille (FR). **NGUYEN, Catherine** [FR/FR]; 8 Boulevard de la Kabylie, F-13016 Marseille (FR). **VIENS, Patrice** [FR/FR]; La Palmeraie Borely, Entrée A, 70, avenue d'Haïfa, F-13008 Marseille (FR). **FERT, Vincent** [FR/FR]; 971, chemin du Tardinaou, F-13190 Allauch (FR).

(74) Agents: **BREESE, Pierre** et al.; Breese-Majerowicz, 3, avenue de l'Opéra, F-75001 Paris (FR).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/46467 A2

(54) Title: GENE EXPRESSION PROFILING OF PRIMARY BREAST CARCINOMAS USING ARRAYS OF CANDIDATE GENES

(57) Abstract: The invention relates to a polynucleotide library useful in the molecular characterization of a carcinoma, the library including a pool of polynucleotide sequences of subsequences thereof wherein the sequences of subsequences are overpressed in tumor cells, further wherein the sequences of subsequences correspond substantially to any of the polynucleotide sequences set forth in any of SEQ ID NOS: 1- 468 or the complement thereof. The invention relates also to polynucleotides arrays useful to differentiate tumor cells from normal cells comprising combinations of selected immobilized polynucleotide sequences sets.

Gene symbol	SET No	Name	Seq3'	Seq5'	Ref
		(autotaxin)			
RELA	19	v-rel avian reticuloendotheliosis viral oncogene homolog a (nuclear factor of kappa light polypeptide gene enhancer in b-cells 3 (p65))	SEQ ID No:42		SEQ ID No:43
ITK	20	il2-inducible t-cell kinase		SEQ ID No:44	SEQ ID No:45
TNXB	21	tenascin xb		SEQ ID No:46	SEQ ID No:47
CSF1	22	colony stimulating factor 1 (macrophage)	SEQ ID No:48	SEQ ID No:49	SEQ ID No:50
VIL2	23	villin 2 (ezrin)	SEQ ID No:51	SEQ ID No:52	SEQ ID No:53
APC	24	adenomatosis polyposis coli	SEQ ID No:54	SEQ ID No:55	SEQ ID No:56
MUC1	25	mucin 1, transmembrane		SEQ ID No:57	SEQ ID No:58
IGF2	26	insulin-like growth factor 2 (somatomedin a)	SEQ ID No:59	SEQ ID No:60	SEQ ID No:61
EMR1	27	egf-like module containing, mucin-like, hormone receptor-like sequence 1	SEQ ID No:62	SEQ ID No:63	SEQ ID No:64
KIAA0427	28	kiaa0427 gene product	SEQ ID No:65	SEQ ID No:66	SEQ ID No:67
SYK	29	spleen tyrosine kinase	SEQ ID No:68	SEQ ID No:69	SEQ ID No:70
IL7R	30	interleukin 7 receptor		SEQ ID No:71	SEQ ID No:72
MYC	31	v-myc avian myelocytomatosis viral oncogene homolog	SEQ ID No:73	SEQ ID No:74	SEQ ID No:75
GATA3	32	gata-binding protein 3	SEQ ID No:76	SEQ ID No:77	SEQ ID No:78
GRB7	33	growth factor receptor-bound protein 7	SEQ ID No:79	SEQ ID No:80	SEQ ID No:81
TOP2B	34	topoisomerase (dna) ii beta (180kd)		SEQ ID No:82	SEQ ID No:83
CASP4	35	caspase 4, apoptosis-related cysteine protease	SEQ ID No:84		SEQ ID No:85
TIMP2	36	tissue inhibitor of metalloproteinase 2		SEQ ID No:86	SEQ ID No:87
DDT	37	d-dopachrome tautomerase	SEQ ID No:88	SEQ ID No:89	SEQ ID No:90
PRL	38	prolactin	SEQ ID	SEQ ID	SEQ ID

The invention concerns also a method for detecting differentially expressed polynucleotide sequences which are correlated with a cancer, said method comprising:

a) obtaining a polynucleotide sample from a patient; and

b) reacting the sample polynucleotide obtained in step (a) with a probe immobilized on a solid support wherein said probe comprises any of the polynucleotide sequences of the libraries previously described or an expression product encoded by any of the polynucleotide sequences of said libraries and

c) detecting the reaction product of step (b).

The invention relates also to a such method for detecting differentially expressed polynucleotide sequences of the invention wherein the amount of reaction product of step (c) is compared to a control sample.

Preferably the polynucleotide sample isolated for, the sample is RNA or mRNA.

Preferably the polynucleotide sample is cDNA obtained by reverse transcription of the mRNA.

In a prefered embodiment the method for detecting differentially expressed polynucleotide sequences, the step (b) comprises a hybridization of the sample RNA with the labeled probe.

The method for detecting differentially expressed polynucleotide sequences is used for detecting, diagnosing, staging, monitoring, prognosticating, preventing or treating conditions associated with cancer, and namelly breast cancer.

The method for detecting differentially expressed polynucleotide sequences is particular useful wherein the product encoded by any of the polynucleotide sequences or

Preferably, the polynucleotide array useful to classify good and poor prognosis primary breast tumors bears any combination of at least one polynucleotide sequence selected among those included in each one of predefined polynucleotide sequences sets defined in table 9A together with at least one polynucleotide sequence selected among those included in each one of predefined polynucleotide sequences sets defined in table 9B.

The present invention concerns also a method for detecting differentially expressed polynucleotide sequences that are correlated with a cancer, said method comprising:

a) obtaining a polynucleotide sample from a patient; and

b) reacting the sample polynucleotide obtained in step (a) with a probe immobilized on a solid support wherein said probe comprises any of the polynucleotide sequences of the libraries previously defined or an expression product encoded by any of the polynucleotide sequences of the libraries previously defined

c) detecting the reaction product of step (b).

Preferably, the polynucléotide sample obtained at step (a) is labeled before its reaction at step (b) with the probe immobilized on a solid support.

The label of the polynucléotide sample is selected from the group consisting of radioactive, colorimetric, enzymatic, molecular amplification, bioluminescent or fluorescent.

In a particular embodiment the reaction product of step (c) is quantified by further comparison of said reaction product to a control sample.

In a first embodiment, the polynucleotide sample isolated from the patient and obtained at step (a) is either RNA or mRNA.

In another embodiment the polynucleotide sample isolated from the patient is cDNA is obtained by reverse transcription of the mRNA.

Preferably the reaction step (b) of the method for detecting differentially expressed polynucleotide sequences comprises a hybridization of the sample RNA issued from patient with the probe.

Preferably the sample RNA is labeled before hybridization with the probe and the label is selected from the group consisting of radioactive, colorimetric, enzymatic, molecular amplification, bioluminescent or fluorescent.

This method for detecting differentially expressed polynucleotide sequences is particularly useful for detecting, diagnosing, staging, monitoring, prognosticating, preventing or treating conditions associated with cancer, and particularly breast cancer.

The method for detecting differentially expressed polynucleotide sequences is also particularly useful when the product encoded by any of the polynucleotide sequences or subsequences set is involved in a receptor-ligand reaction on which detection is based.

The present invention is also related with a method for screening an anti-tumor agent comprising the method the above-depicted method for detecting differentially expressed polynucleotide sequences wherein the sample has been treated with the anti-tumor agent to be screened.

In a particular embodiment the method for screening an anti-tumor agent comprises detecting polynucleotide sequences reacting with at least one library of polynucleotides or polynucleotide sequences set as previously defined or of products encoded by said library in a sample obtained from a patient.

The invention is illustrated by examples detailed below related to particular experimental results obtained with selected libraries of polypeptides useful to identify and distinguish tumor samples from normal ones.

Tumor samples and RNA extraction

To avoid any bias of selection as to the type and size of the tumors, the RNAs to be tested were prepared from unselected samples. Samples of primary invasive breast carcinomas were collected from 34 patients undergoing surgery at the Institute Paoli-Calmette. After surgical resection, the tumors were macrodissected: a section was taken for the pathologist's diagnosis and an adjacent piece was quickly frozen in liquid nitrogen for molecular analyses. The median age of patients at the time of diagnosis was 55 years (range 39, 83) and most of them were post-menopausal. Tumors were classified according to the WHO histological typing of breast tumors in: 29 ductal carcinomas, 2 lobular carcinomas, 1 mixed ductal and lobular carcinoma, and 2 medullar carcinomas. They had various sizes, inferior or equal to 20 mm (n = 13), between 20 and 50 mm (n = 18) or superior to 50 mm (n = 3), axillary's lymph node status (negative: 19 tumors, positive: 15 tumors), SBR grading (I: 3 tumors, II: 20 tumors, III: 10 tumors, not evaluable: 1 tumor), and estrogen receptor status (ER) evaluated by

immunohistochemical assay (23 ER-positive, 11 ER-negative). ER positivity cutoff value was 10%. Adjuvant treatment with radiotherapy and when necessary multi-agent anthracyclin-based chemotherapy (n = 16) was given to patients according to local practice.

Total RNA was extracted from tumor samples by standard methods (43). Total RNA from normal breast tissue was obtained from Clontech (Palo Alto, CA): RNA was isolated from 8 tissue specimens from Caucasian females, age range 23 - 47. RNA integrity was controlled by denaturing formaldehyde agarose gel electrophoresis and Northern blots using a 28S-specific oligonucleotide.

cDNA arrays preparation

Gene expression was analyzed by hybridization of arrays with radioactive probes. The arrays contained PCR products of 5 control clones, and 180 IMAGE human cDNA clones selected with practical criteria (3' sequence of mRNA, same cloning vector, host bacteria and insert size). This represented 176 genes (4 genes were represented by 2 different clones): 121 with proven or putative implication in cancer and 55 implicated in immune reactions (the list is available on the web site: http:/tagc.univ-mrs.fr/pub/Cancer/). Their identity was verified by 5' tag-sequencing of plasmid DNA and comparison with sequences in the EST (dbEST) and nucleotide (GenBank) databases at the NCBI. Identity was confirmed for all but 14 clones without significant gene similarity, which were referenced by their GenBank accession number. The control clones were: Arabidopsis thaliana cytochrome c554 gene (used for hybridization signal normalization), 3 poly(A) sequences of different sizes and the vector pT7T3D (negative controls).

PCR amplification, purification and robotical spotting of PCR products onto Hybond-N+ membranes (Amersham) were done according to described protocols (4). All PCR products were spotted in duplicate. For normalization purpose, the c554 gene was spotted 96-fold scattered over the whole membrane.

cDNA array hybridizations

Hybridizations were done successively with a vector oligonucleotide (to precisely determine the amount of target DNA accessible to hybridization in each spot), then after stripping of vector probe, with complex probes made from the RNAs (4). Each complex probe was hybridized to a distinct filter. Probes were prepared from total RNA with an excess of oligo(dT25) to saturate the poly(A) tails of the messengers, and to insure that the reverse transcribed product did not contain long poly(T) sequences. A precise amount of c554 mRNA was added to the total RNA before labeling to allow normalization of the data.

Five ng of total RNA (~100ng of mRNA) from tissue samples were used for each labeling. Probe preparation and hybridization of the membranes were done according to known procedures (http:/tagc.univ-mrs.fr/pub/Cancer/).

Hybridization was done in excess of target (~15 ng of DNA in each spot) and binding of cDNAs to the targets was linear and proportional to the quantity of cDNA in the probe.

Detection and quantification of cDNA array hybridization signals

Quantitative data were obtained using an imaging plate device. Hybridization signal detection with a FUJI BAS 1500 machine and quantification with the HDG Analyzer

software (Genomic Solutions, Ann Arbor, MI) were done as previously described (http:/tagc.univ-mrs.fr/pub/Cancer/). Quantification was done by integrating all spot pixel intensities and substracting a spot background value determined in the neighboring area. Spots were located with a LaPlacian transformation. Spot background level was the median intensity of all the pixels present in a small window centered on the spot and which were not part of any spot (44). Quantified data were normalized in three steps and expressed as absolute gene expression levels (i.e. in percentage of abundance of individual mRNA with respect to mRNA within the sample), as described (4).

Array data analysis

Before analysis of the results, the reproducibility of the experiments was verified by comparing duplicate spots, or one hybridization with the same probe on two independent arrays, or two independent hybridizations with probes prepared from the same RNA. In every case, the results showed good reproducibility with respective correlation coefficients of 0.95, 0.98 and 0.98 (data not shown). Moreover, genes represented by two different clones on the array, such as CDK4 or ETV5, displayed similar expression profiles for the two clones in all samples. This reproducibility was sufficient enough to consider a 2-fold expression difference as significantly differential.

For graphical representation, data were displayed as absolute expression levels (Fig. 2a). For better visualization of clustering, results were log-transformed and displayed as relative values median-centered in each row and in each column (Fig. 2b). Hierarchical clustering was applied to the tissue samples and the genes using the Cluster program developed by Eisen (45) (average linkage clustering

using Pearson correlation as similarity metric). Results in Figs. 2 and 3 were displayed with the TreeView program (45).

Subsequent analysis was done using Excel software (Microsoft) and statistical analyses with the SPSS software. Metastasis-free survival and overall survival were measured from diagnosis until the first metastatic relapse or death respectively. They were estimated with the Kaplan-Meier method and compared between groups with the Log-Rank test. Correlations of gene pairs based on expression profiles were measured with the correlation coefficient r. The search for genes with expression levels correlated with tumor parameters was done in several successive steps.

First, genes were detected by comparing their median expression level in the two subgroups of tumors discordant according to the parameter of interest. The median values rather than the mean values were used because of the high variability of the expression levels for many genes, resulting in a standard deviation of expression level similar or superior to the mean value and making comparisons with means impossible. Second, these detected genes were inspected visually on graphics, and finally, an appropriate statistical analysis was applied to those that were convincing to validate the correlation. Comparison of GATA3 expression between ER-positive tumors and ER-negative tumors was validated using a Mann-Witney test. Correlation coefficients were used to compare the gene expression levels to the number of axillary nodes involved.

Northern blot analysis

Seventy-nine breast tumors, including 22 of the 34 tested on the arrays, were analyzed for GATA3 expression by Northern blot hybridization. RNA extraction from tumor samples and Northern blots were done as previously described

CLAIMS

1. A polynucleotide library useful in the molecular characterization of a carcinoma, said library comprising a pool of polynucleotide sequences or subsequences thereof wherein said sequences or subsequences are either underexpressed or overpressed in tumor cells, further wherein said sequences or subsequences correspond substantially to any of the polynucleotide sequences set forth in any of SEQ ID Nos: 1 - 468 or the complement thereof.

2. A polynucleotide library according to Claim 1 wherein said polynucleotide sequences or subsequences thereof of said pool correspond to any combination of at least one polynucleotide selected among those included in anyone of the following predefined sets :

SET 1: (SEQ ID No:1; SEQ ID No:2); SET 2: (SEQ ID No:3; SEQ ID No:4); SET 3: (SEQ ID No:5; SEQ ID No:6); SET 4: (SEQ ID No:7;SEQ ID No:8); SET 5: (SEQ ID No:9; SEQ ID No:10); SET 6: (SEQ ID No:11: SEQ ID No:12); SET 7: (SEQ ID No:13; SEQ ID No:14;SEQ ID No:15); SET 8: (SEQ ID No:16); SET 9: (SEQ ID No:17; SEQ ID No:18; SEQ ID No:19); SET 10: (SEQ ID No:20; SEQ ID No:21); SET 11: (SEQ ID No:22; SEQ ID No:23; SEQ ID No:24); SET 12: (SEQ ID No:25; SEQ ID No:26); SET 13: (SEQ ID No:27; SEQ ID No:28; SEQ ID No:29); SET 14: (SEQ ID No:30; SEQ ID No:31); SET 15: (SEQ ID No:32; SEQ ID No:33; SEQ ID No:34)) ; SET 16 : (SEQ ID No:35) ; SET 17 : (SEQ ID No:36; SEQ ID No:37; SEQ ID No:38) ; SET 18 : (SEQ ID No:39; SEQ ID No:40; SEQ ID No:41) ; SET 19 : (SEQ ID No:42; SEQ ID No:43) ; SET 20 : (SEQ ID No:44; SEQ ID No:45) ; SET 21 : (SEQ ID No:46; SEQ ID No:47) ; SET 22 : (SEQ ID No:48; SEQ ID No:49; SEQ ID No:50) ; SET 23 : (SEQ ID No:51; SEQ ID No:52; SEQ ID No:53) ; SET 24: (SEQ ID No:54; SEQ ID No:55; SEQ ID No:56) ; SET 25: (SEQ ID No:57; SEQ ID No:58) ; SET 26: (SEQ ID No:59; SEQ ID No:60; SEQ ID No:61) ; SET 27: (SEQ ID No:62; SEQ ID No:63; SEQ ID No:64) ; SET 28: (SEQ ID No:65; SEQ ID No:66; SEQ ID No:67) ;

SET 29: (SEQ ID No:68; SEQ ID No:69; SEQ ID No:70) ; SET 30: (SEQ ID No:71; SEQ ID No:72) ; SET 31 : (SEQ ID No:73; SEQ ID No:74; SEQ ID No:75) ; SET 32 : (SEQ ID No:76; SEQ ID No:77; SEQ ID No:78) ; SET 33 : (SEQ ID No:79; SEQ ID No:80; SEQ ID No:81) ; SET 34: (SEQ ID No:82; SEQ ID No:83) ; SET 35: (SEQ ID No:84; SEQ ID No:85) ; SET 36: (SEQ ID No:86; SEQ ID No:87) ; SET 37: (SEQ ID No:88; SEQ ID No:89; SEQ ID No:90) ; SET 38: (SEQ ID No:91; SEQ ID No:92; SEQ ID No:93) ; SET 39: (SEQ ID No:94; SEQ ID No:95; SEQ ID No:96) ; SET 40: (SEQ ID No:97; SEQ ID No:98; SEQ ID No:99) ; SET 41: (SEQ ID No:100; SEQ ID No:101; SEQ ID No:78) ; SET 42: (SEQ ID No:102; SEQ ID No:103) ; SET 43: (SEQ ID No:104; SEQ ID No:105) ; SET 44: (SEQ ID No:106; SEQ ID No:107; SEQ ID No:108) ; SET 45: (SEQ ID No:109; SEQ ID No:110) ; SET 46: (SEQ ID No:111; SEQ ID No:112; SEQ ID No:113) ; SET 47: (SEQ ID No:114) ; SET 48: (SEQ ID No:115; SEQ ID No:116; SEQ ID No:117) ; SET 49: (SEQ ID No:118; SEQ ID No:119) ; SET 50: (SEQ ID No:120; SEQ ID No:121) ; SET 51: (SEQ ID No:122; SEQ ID No:78) ; SET 52: (SEQ ID No:123; SEQ ID No:124; SEQ ID No:125) ; SET 53: (SEQ ID No:126; SEQ ID No:127; SEQ ID No:128) ; SET 54: (SEQ ID No:129; SEQ ID No:130) ; SET 55: (SEQ ID No:131; SEQ ID No:132) ; SET 56: (SEQ ID No:133; SEQ ID No:134) ; SET 57: (SEQ ID No:135; SEQ ID No:136; SEQ ID No:137) ; SET 58: (SEQ ID No:138; SEQ ID No:139; SEQ ID No:140) ; SET 59: (SEQ ID No:141; SEQ ID No:142; SEQ ID No:143) ; SET 60: (SEQ ID No:144; SEQ ID No:145; SEQ ID No:146) ; SET 61: (SEQ ID No:147; SEQ ID No:148; SEQ ID No:149) ; SET 62: (SEQ ID No:150; SEQ ID No:151; SEQ ID No:152) ; SET 63: (SEQ ID No:153; SEQ ID No:154; SEQ ID No:155) ; SET 64: (SEQ ID No:156; SEQ ID No:157; SEQ ID No:158) ; SET 65: (SEQ ID No:159; SEQ ID No:160; SEQ ID No:161) ; SET 66: (SEQ ID No:162; SEQ ID No:163) ; SET 67: (SEQ ID No:164; SEQ ID No:165) ; SET 68: (SEQ ID No:166; SEQ ID No:167; SEQ ID No:152) ; SET 69: (SEQ ID No:168; SEQ ID No:169; SEQ ID No:170) ; SET 70: (SEQ ID No:171; SEQ ID No:172) ; SET 71: (SEQ ID No:173; SEQ ID No:174; SEQ ID No:175) ; SET 72: (SEQ ID No:176; SEQ ID No:177) ; SET 73: (SEQ ID No:178; SEQ ID No:179) ; SET 74: (SEQ ID No:180; SEQ ID No:181; SEQ ID No:182) ; SET 75: (SEQ ID No:183; SEQ ID No:184) ; SET 76: (SEQ ID No:185) ; SET 77: (SEQ ID No:186) ; SET 78: (SEQ ID No:187; SEQ ID No:188) ; SET 79: (SEQ ID No:189;

SEQ ID No:190; SEQ ID No:191) ; SET 80: (SEQ ID No:192; SEQ ID No:193) ; SET 81: (SEQ ID No:194; SEQ ID No:195) ; SET 82: (SEQ ID No:196; SEQ ID No:197; SEQ ID No:198) ; SET 83: (SEQ ID No:199; SEQ ID No:200) ; SET 84: (SEQ ID No:201; SEQ ID No:202; SEQ ID No:203) ; SET 85: (SEQ ID No:204; SEQ ID No:205) ; SET 86: (SEQ ID No:206; SEQ ID No:207) ; SET 87: (SEQ ID No:208; SEQ ID No:209) ; SET 88: (SEQ ID No:210; SEQ ID No:211) ; SET 89: (SEQ ID No:212; SEQ ID No:213) ; SET 90: (SEQ ID No:214; SEQ ID No:215) ; SET 91: (SEQ ID No:216; SEQ ID No:217) ; SET 92: (SEQ ID No:218; SEQ ID No:219; SEQ ID No:220) ; SET 93: (SEQ ID No:221; SEQ ID No:222) ; SET 94: (SEQ ID No:223; SEQ ID No:224; SEQ ID No:225) ; SET 95: (SEQ ID No:226; SEQ ID No:227) ; SET 96: (SEQ ID No:228; SEQ ID No:229) ; SET 97: (SEQ ID No:230; SEQ ID No:231; SEQ ID No:232) ; SET 98: (SEQ ID No:233; SEQ ID No:234) ; SET 99: (SEQ ID No:235; SEQ ID No:236; SEQ ID No:237) ; SET 100: (SEQ ID No:238; SEQ ID No:239) ; SET 101: (SEQ ID No:240; SEQ ID No:241) ; SET 102: (SEQ ID No:242; SEQ ID No:243; SEQ ID No:244) ; SET 103: (SEQ ID No:245; SEQ ID No:246; SEQ ID No:247) ; SET 104: (SEQ ID No:248; SEQ ID No:249) ; SET 105: (SEQ ID No:250; SEQ ID No:251; SEQ ID No:252) ; SET 106: (SEQ ID No:253; SEQ ID No:254) ; SET 107: (SEQ ID No:255; SEQ ID No:256) ; SET 108: (SEQ ID No:257; SEQ ID No:258) ; SET 109: (SEQ ID No:259; SEQ ID No:260; SEQ ID No:261) ; SET 110: (SEQ ID No:262; SEQ ID No:200) ; SET 111: (SEQ ID No:263; SEQ ID No:264) ; SET 112: (SEQ ID No:265; SEQ ID No:266) ; SET 113: (SEQ ID No:267; SEQ ID No:268) ; SET 114: (SEQ ID No:269; SEQ ID No:270) ; SET 115: (SEQ ID No:271; SEQ ID No:272) ; SET 116: (SEQ ID No:273; SEQ ID No:274) ; SET 117: (SEQ ID No:275; SEQ ID No:276) ; SET 118: (SEQ ID No:277; SEQ ID No:278) ; SET 119: (SEQ ID No:279; SEQ ID No:280; SEQ ID No:281) ; SET 120: (SEQ ID No:282; SEQ ID No:283; SEQ ID No:276) ; SET 121: (SEQ ID No:284; SEQ ID No:285) ; SET 122: (SEQ ID No:286; SEQ ID No:287; SEQ ID No:288) ; SET 123: (SEQ ID No:289; SEQ ID No:290) ; SET 124: (SEQ ID No:291; SEQ ID No:292) ; SET 125: (SEQ ID No:293; SEQ ID No:294; SEQ ID No:295) ; SET 126: (SEQ ID No:296; SEQ ID No:297) ; SET 127: (SEQ ID No:298; SEQ ID No:299; SEQ ID No:300) ; SET 128: (SEQ ID No:301; SEQ ID No:302; SEQ ID No:288) ; SET 129: (SEQ ID No:303; SEQ ID No:304) ; SET 130: (SEQ ID No:305; SEQ ID No:306;

SEQ ID No:307) ; SET 131: (SEQ ID No:308; SEQ ID No:309; SEQ ID No:310) ; SET 132: (SEQ ID No:311; SEQ ID No:312; SEQ ID No:313) ; SET 133: (SEQ ID No:314; SEQ ID No:315; SEQ ID No:316) ; SET 134: (SEQ ID No:317; SEQ ID No:318) ; SET 135: (SEQ ID No:319; SEQ ID No:320; SEQ ID No:321) ; SET 136: (SEQ ID No:322; SEQ ID No:323) ; SET 137: (SEQ ID No:324; SEQ ID No:325) ; SET 138: (SEQ ID No:326; SEQ ID No:327; SEQ ID No:328) ; SET 139: (SEQ ID No:329; SEQ ID No:330) ; SET 140: (SEQ ID No:331; SEQ ID No:332; SEQ ID No:333) ; SET 141: (SEQ ID No:334; SEQ ID No:335; SEQ ID No:336) ; SET 142: (SEQ ID No:337; SEQ ID No:338; SEQ ID No:117) ; SET 143: (SEQ ID No:339; SEQ ID No:340; SEQ ID No:341) ; SET 144: (SEQ ID No:342; SEQ ID No:343; SEQ ID No:344) ; SET 145: (SEQ ID No:345; SEQ ID No:346) ; SET 146: (SEQ ID No:347; SEQ ID No:348; SEQ ID No:349) ; SET 147: (SEQ ID No:350; SEQ ID No:351) ; SET 148: (SEQ ID No:352; SEQ ID No:353) ; SET 149: (SEQ ID No:354; SEQ ID No:355) ; SET 150: (SEQ ID No:356; SEQ ID No:357) ; SET 151: (SEQ ID No:358; SEQ ID No:359; SEQ ID No:360) ; SET 152: (SEQ ID No:361; SEQ ID No:31) ; SET 153: (SEQ ID No:362; SEQ ID No:363; SEQ ID No:364) ; SET 154: (SEQ ID No:365; SEQ ID No:366; SEQ ID No:367) ; SET 155: (SEQ ID No:368; SEQ ID No:369; SEQ ID No:300) ; SET 156: (SEQ ID No:370; SEQ ID No:371) ; SET 157: (SEQ ID No:372; SEQ ID No:373; SEQ ID No:108) ; SET 158: (SEQ ID No:374; SEQ ID No:375; SEQ ID No:376) ; SET 159: (SEQ ID No:377; SEQ ID No:378; SEQ ID No:379) ; SET 160: (SEQ ID No:380; SEQ ID No:381) ; SET 161: (SEQ ID No:382; SEQ ID No:383; SEQ ID No:384) ; SET 162: (SEQ ID No:385; SEQ ID No:386; SEQ ID No:387) ; SET 163: (SEQ ID No:388; SEQ ID No:389; SEQ ID No:390) ; SET 164: (SEQ ID No:391; SEQ ID No:392; SEQ ID No:393) ; SET 165: (SEQ ID No:394; SEQ ID No:395) ; SET 166: (SEQ ID No:396; SEQ ID No:397; SEQ ID No:398) ; SET 167: (SEQ ID No:399; SEQ ID No:400; SEQ ID No:117) ; SET 168: (SEQ ID No:401) ; SET 169: (SEQ ID No:402; SEQ ID No:403) ; SET 170: (SEQ ID No:404; SEQ ID No:405; SEQ ID No:318) ; SET 171: (SEQ ID No:406; SEQ ID No:407; SEQ ID No:408) ; SET 172: (SEQ ID No:409; SEQ ID No:410; SEQ ID No:411) ; SET 173: (SEQ ID No:412; SEQ ID No:413) ; SET 174: (SEQ ID No:414; SEQ ID No:415; SEQ ID No:416) ; SET 175: (SEQ ID No:417; SEQ ID No:418; SEQ ID No:419) ; SET 176: (SEQ ID No:420; SEQ ID No:421; SEQ ID No:422) ; SET 177: (SEQ ID No:423;

SEQ ID No:424; SEQ ID No:425) ; SET 178: (SEQ ID No:426; SEQ ID No:427; SEQ ID No:428) ; SET 179: (SEQ ID No:429; SEQ ID No:408) ; SET 180: (SEQ ID No:430) ; SET 181: (SEQ ID No:431) ; SET 182: (SEQ ID No:432) ; SET 183: (SEQ ID No:433; SEQ ID No:434) ; SET 184: (SEQ ID No:435; SEQ ID No:436) ; SET 185: (SEQ ID No:437) ; SET 186: (SEQ ID No:438; SEQ ID No:439) ; SET 187: (SEQ ID No:440; SEQ ID No:441) ; SET 188: (SEQ ID No:442) ; SET 189: (SEQ ID No:444) ; SET 190: (SEQ ID No:445) ; SET 191 (SEQ ID No:446 ; SEQ ID No:447) ; SET 192: (SEQ ID No:448) ; SET 193: (SEQ ID No:449) ; SET 194: (SEQ ID No:450): SET 195: (SEQ ID No:451) ; SET 196: (SEQ ID No:452) ; SET 197: (SEQ ID No:453) ; SET 198: (SEQ ID No:454) ; SET 199: (SEQ ID No:455) ; SET 200: (SEQ ID No:456) ; SET 201: (SEQ ID No:457) ; SET 202: (SEQ ID No:458) ; SET 203: (SEQ ID No:459) ; SET 204: (SEQ ID No:460) ; SET 205: (SEQ ID No:461) ; SET 206: (SEQ ID No:462) ; SET 207: (SEQ ID No:463) ; SET 208: (SEQ ID No:464) ; SET 209: (SEQ ID No:465) ; SET 210: (SEQ ID No:466) ; SET 211: (SEQ ID No:467) ; SET 212: (SEQ ID No:468)

3. A polynucleotide library according to Claim 2 wherein said polynucleotide sequences or subsequences thereof of said pool correspond to any combination of at least one polynucleotide selected among those included in at least 50%, preferably 75% and more preferably 100% of the predefined sets.

4. A library according to anyone Claim 1 or 2 wherein the pool of polynucleotide sequences or subsequences correspond substantially to any combination of at least one polynucleotide sequence selected among those included in each one of predefined polynucleotide sequences sets comprising:

SET 1: (SEQ ID No:1 ; SEQ ID No:2) ; SET 4: (SEQ ID No:7 ; SEQ ID No:8) ; SET 18: (SEQ ID No:39 ; SEQ ID No:40 ; SEQ ID No:41) ; SET 21: (SEQ ID No:46 ; SEQ ID No:47) ; SET 24: (SEQ ID No:54 ; SEQ ID No:55 ; SEQ ID No:56) ; SET 32: (SEQ ID No:76 ; SEQ ID No:77 ; SEQ ID No:78) ; SET 38: (SEQ ID No:91 ; SEQ ID

26. A polynucleotide library according to Claim 25 wherein the support is selected from the group comprising a nylon membrane, nitrocellulose membrane, glass slide, glass beads, membranes on glass support or a silicon chip.

27. A polynucleotide array useful for prognosis or diagnostic of tumor comprising an immobilized polynucletide library according to Claims 1 to 3.

28. A polynucleotide array useful to differentiate a normal cell from a cancer cell comprising any combination of immobilized polynucletide sequences sets according to claims 4 to 7.

29. A polynucleotide array useful to detect a hormone sensitive tumor cell comprising any combination of immobilized polynucletide sequences sets according to claims 8 to 11.

30. A polynucleotide array useful to differentiate a tumor with lymph nodes from a tumor without lymph nodes comprising any combination of immobilized polynucletide sequences sets according to claims 12 to 15.

31. A polynucleotide array useful to differentiate antracycline-sensitive tumors from antracycline-insensitive tumors comprising any combination of immobilized polynucletide sequences sets according to claims 16 to 19.

32. A polynucleotide array useful to classify good and poor prognosis primary breast tumors comprising any

Figure 2

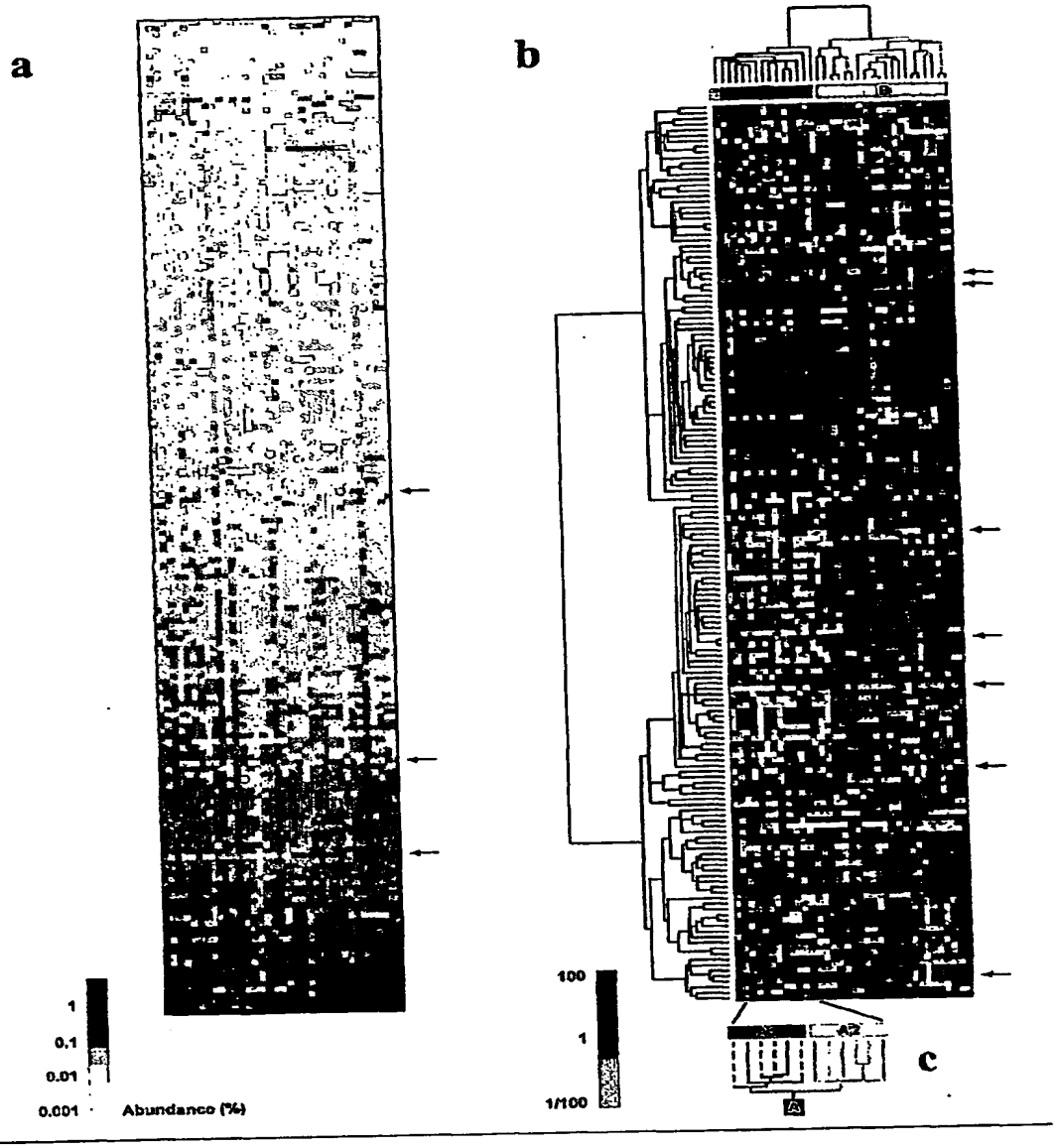

Figure 3

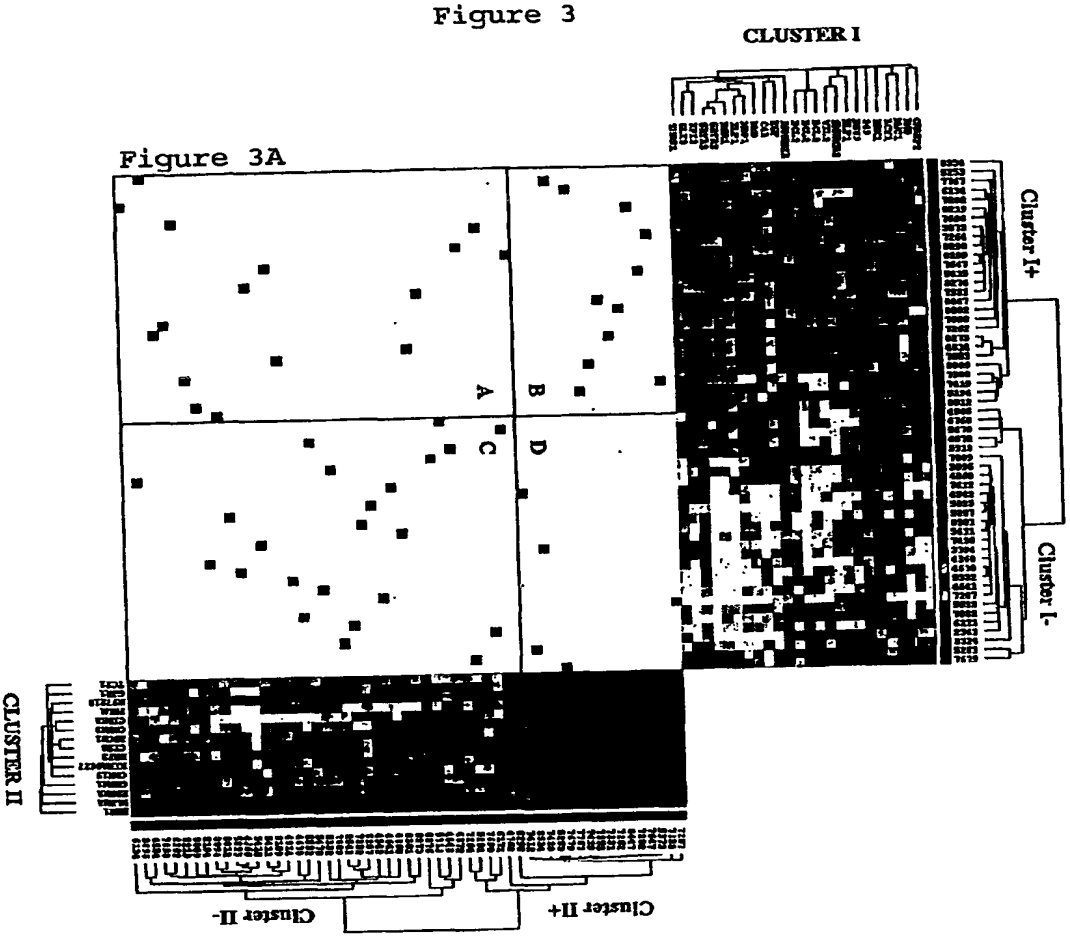

Figure 3C

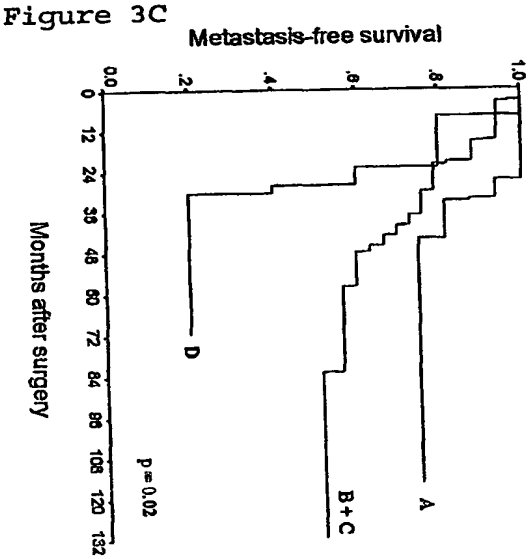

Figure 3B